

Эконометрика

Методические указания и задания к типовой работе
для студентов 1-го курса заочного отделения
специальности 080100 «Экономика» (Э-4, Э-3(СПО))

Контрольная работа состоит из задачи приведенных в конце методических указаний. Данные для каждого варианта определяться двумя параметрами, указанными в конце.

1. Определение эконометрики

Эконометрика – быстроразвивающаяся отрасль науки, цель которой состоит в том, чтобы придать количественные меры экономическим отношениям.

Термин «эконометрика» был впервые введен бухгалтером П. Цьемпой (Австро-Венгрия, 1910 г.). Цьемпа считал, что если к данным бухгалтерского учета применить методы алгебры и геометрии, то будет получено новое, более глубокое представление о результатах хозяйственной деятельности. Это употребление термина, как и сама концепция, не прижилось, но название «эконометрика» оказалось весьма удачным для определения нового направления в экономической науке, которое выделилось в 1930 г.

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. «метрон»). Таким образом, сам термин подчеркивает специфику, содержание эконометрики как науки: количественное выражение тех связей и соотношений, которые раскрыты и обоснованы экономической теорией. И. Шумпетер (1883–1950), один из первых сторонников выделения этой новой дисциплины, полагал, что в соответствии со своим назначением эта дисциплина должна называться «экономометрика». Советский ученый А.Л. Вайнштейн (1892–1970) считал, что название настоящей науки основывается на греческом слове *метрия* (геометрия, планиметрия и т.д.), соответственно по аналогии – эконометрия. Однако в мировой науке общеупотребимым стал термин «эконометрика». В любом случае, какой бы мы термин ни выбрали, эконометрика является наукой об измерении и анализе экономических явлений.

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эта наука возникла

в результате взаимодействия и объединения в особый «сплав» трех компонент: экономической теории, статистических и математических методов. Впоследствии к ним присоединилось развитие вычислительной техники как условие развития эконометрики.

В журнале «Эконометрика», основанном в 1933 г. Р. Фришем (1895–1973), он дал следующее определение эконометрики: «Эконометрика – это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек – статистика, экономическая теория и математика – необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это – единство всех трех составляющих. И это единство образует эконометрику».

Таким образом, эконометрика – это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов.

2. Парная регрессия и корреляция

2.1. Теоретическая справка

Парная (простая) линейная регрессия представляет собой модель, где среднее значение зависимой (объясняемой) переменной рассматривается как функция одной независимой (объясняющей) переменной x , т.е. это модель вида:

$$\hat{y}_x = f(x). \quad (2.1)$$

Так же y называют результативным признаком, а x признаком-фактором. Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости.

Практически в каждом отдельном случае величина y складывается из двух слагаемых:

$$y = \hat{y}_x + \varepsilon, \quad (2.2)$$

где y – фактическое значение результативного признака; \hat{y}_x – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина ε называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

Различают *линейные* и *нелинейные* регрессии.

Линейная регрессия: $y = a + b \cdot x + \varepsilon$.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но

линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам. Например:

регрессии, *нелинейные по объясняющим переменным*:

- полиномы разных степеней $y = a + b_1 \cdot x + b_2 \cdot x^2 + \dots + b_n \cdot x^n + \varepsilon$;
- равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$;

регрессии, *нелинейные по оцениваемым параметрам*:

- степенная $y = a \cdot x^b \cdot \varepsilon$;
- показательная $y = a \cdot b^x \cdot \varepsilon$;
- экспоненциальная $y = e^{a+bx+\varepsilon}$.

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют *метод наименьших квадратов (МНК)*. МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических \hat{y}_x минимальна, т.е.

$$\sum (y - \hat{y}_x)^2 \rightarrow \min \quad (2.3)$$

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b :

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum xy. \end{cases} \quad (2.4)$$

Можно воспользоваться готовыми формулами, которые вытекают непосредственно из решения этой системы:

$$a = \bar{y} - b \cdot \bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2}, \quad (2.5)$$

где $\text{cov}(x, y) = \overline{y \cdot x} - \bar{y} \cdot \bar{x}$ – ковариация признаков x и y , $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – дисперсия признака x и

$$\bar{x} = \frac{1}{n} \sum x, \quad \bar{y} = \frac{1}{n} \sum y, \quad \overline{y \cdot x} = \frac{1}{n} \sum y \cdot x, \quad \overline{x^2} = \frac{1}{n} \sum x^2.$$

(Ковариация – числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений этих случайных величин от их математических ожиданий. Дисперсия – характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Математическое ожидание – сумма произведений значений случайной величины на соответствующие вероятности.)

Тесноту связи изучаемых явлений оценивает *линейный коэффициент парной корреляции* r_{xy} для линейной регрессии ($-1 \leq r_{xy} \leq 1$):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad (2.6)$$

и *индекс корреляции* ρ_{xy} – для нелинейной регрессии ($0 \leq \rho_{xy} \leq 1$):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}},$$

где $\sigma_y^2 = \sum (y - \bar{y})^2$ – общая дисперсия результативного признака y ; $\sigma_{\text{ост}}^2 = \sum (y - \hat{y}_x)^2$ – остаточная дисперсия, определяемая исходя из уравнения регрессии $\hat{y}_x = f(x)$.

Оценку качества построенной модели даст коэффициент (индекс) детерминации r_{xy}^2 (для линейной регрессии) либо ρ_{xy}^2 (для нелинейной регрессии), а также средняя ошибка аппроксимации.

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% . \quad (2.7)$$

Допустимый предел значений \bar{A} – не более 10%.

Средний коэффициент эластичности $\bar{\varepsilon}$ показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины при изменении фактора x на 1% от своего среднего значения:

$$\bar{\varepsilon} = f'(x) \frac{\bar{x}}{\bar{y}} . \quad (2.8)$$

После того как найдено уравнение линейной регрессии, проводится *оценка значимости* как уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Оценка значимости уравнения регрессии в целом производится на основе *F-критерия Фишера*, которому предшествует дисперсионный анализ. Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения \bar{y} раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2 ,$$

где $\sum (y - \bar{y})^2$ – общая сумма квадратов отклонений; $\sum (\hat{y}_x - \bar{y})^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма

квадратов отклонений); $\sum (y - \hat{y}_x)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Схема дисперсионного анализа имеет вид, представленный в таблице 1.1 (n – число наблюдений, m – число параметров при переменной x).

Таблица 2.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum (y - \bar{y})^2$	$n - 1$	$S_{\text{общ}}^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$
Факторная	$\sum (\hat{y}_x - \bar{y})^2$	m	$S_{\text{факт}}^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{m}$
Остаточная	$\sum (y - \hat{y}_x)^2$	$n - m - 1$	$S_{\text{ост}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - m - 1}$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду (напомним, что степени свободы – это числа, показывающие количество элементов варьирования, которые могут принимать произвольные значения, не изменяющие заданных характеристик). Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F -критерия Фишера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2}.$$

Фактическое значение F -критерия Фишера сравнивается с табличным значением $F_{\text{табл}}(\alpha; k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F -критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \hat{y}_x)^2} \cdot (n - 2).$$

Величина F -критерия связана с коэффициентом детерминации r_{xy}^2 , и ее можно рассчитать по следующей формуле:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2). \quad (2.9)$$

Для оценки *статистической значимости параметров регрессии и корреляции* рассчитываются *t-критерий Стьюдента* и *доверительные интервалы* каждого из показателей. Оценка значимости коэффициентов регрессии и корреляции с помощью *t-критерия Стьюдента* проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}; \quad t_r = \frac{r_{xy}}{m_r}. \quad (2.10)$$

Стандартные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$\begin{aligned} m_b &= \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_{\text{ост}}^2}{n \cdot \sigma_x^2}}; \\ m_a &= \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n - 2)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_{\text{ост}}^2 \frac{\sum x^2}{n^2 \sigma_x^2}}; \\ m_{r_{xy}} &= \sqrt{\frac{1 - r_{xy}^2}{n - 2}}. \end{aligned} \quad (2.11)$$

Сравнивая фактическое и критическое (табличное) значения t -статистики — $t_{\text{табл}}$ и $t_{\text{факт}}$ — делаем вывод о значимости параметров регрессии и корреляции. Если $t_{\text{табл}} < t_{\text{факт}}$ то параметры a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x . Если $t_{\text{табл}} > t_{\text{факт}}$, то признается случайная природа формирования a , b или r_{xy} .

Для расчета доверительного интервала определяем *предельную ошибку* Δ для каждого показателя:

$$\Delta_a = t_{\text{табл}} m_a, \quad \Delta_b = t_{\text{табл}} m_b.$$

Формулы для расчета *доверительных интервалов* имеют следующий вид:

$$\begin{aligned} \gamma_a &= a \pm \Delta_a; & \gamma_{a_{\min}} &= a - \Delta_a; & \gamma_{a_{\max}} &= a + \Delta_a; \\ \gamma_b &= b \pm \Delta_b; & \gamma_{b_{\min}} &= b - \Delta_b; & \gamma_{b_{\max}} &= b + \Delta_b; \end{aligned}$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

Связь между F -критерием Фишера и t -статистикой Стьюдента выражается равенством

$$|t_r| = |t_b| = \sqrt{F}. \quad (2.12)$$

В прогнозных расчетах по уравнению регрессии определяется *предсказываемое индивидуальное значение* y_0 как точечный прогноз при $x = x_0$, т.е. путем подстановки в линейное уравнение $\hat{y}_x = a + b \cdot x$ соответствующего значения x . Однако точечный прогноз явно нереален, поэтому он дополняется расчетом стандартной ошибки

$$m_{\hat{y}_0} = \sqrt{S_{\text{ост}}^2 \left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2} \right)} = \sqrt{S_{\text{ост}}^2 \left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2} \right)}, \quad (2.13)$$

где $S_{\text{ост}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2}$, и построением *доверительного интервала* прогнозного значения y_0^* :

$$\hat{y}_0 - m_{\hat{y}_0} \cdot t_{\text{табл}} \leq y_0^* \leq \hat{y}_x + m_{\hat{y}_x} \cdot t_{\text{табл}}$$

2.2. Решение типовой задачи

Пример. По территориям региона приводятся данные за 199X г.

Таблица 2.2

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Требуется:

1. Построить линейное уравнение парной регрессии y по x .
2. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F -критерия Фишера и t -критерия Стьюдента.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x , составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.
6. На одном графике отложить исходные данные и теоретическую прямую.

Решение

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу 2.3.

Таблица 2.3

№	x	y	$y \cdot x$	x^2	y^2	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	A_i
1	2	3	4	5	6	7	8	9	10
1	78	133	10374	6084	17689	148,78	-15,78	249,01	11,86
2	82	148	12136	6724	21904	152,46	-4,46	19,89	3,01
3	87	134	11658	7569	17956	157,06	-23,06	531,76	17,21
4	79	154	12166	6241	23716	149,70	4,30	18,49	2,79
5	89	162	14418	7921	26244	158,90	3,10	9,61	1,91
6	106	195	20670	11236	38025	174,54	20,46	418,61	10,49
7	67	139	9313	4489	19321	138,66	0,34	0,12	0,24
8	88	158	13904	7744	24964	157,98	0,02	0,00	0,01
9	73	152	11096	5329	23104	144,18	7,82	61,15	5,14
10	87	162	14094	7569	26244	157,06	4,94	24,40	3,05
11	76	159	12084	5776	25281	146,94	12,06	145,44	7,58
12	115	173	19895	13225	29929	182,82	-9,82	96,43	5,68
Итого	1027	1869	161808	89907	294377	1869,08	-0,08	1574,91	68,97
Среднее значение	85,58	155,75	13484,0	7492,25	24531,4	155,76	—	131,24	5,75
σ	12,97	16,53	—	—	—	—	—	—	—
σ^2	168,31	273,34	—	—	—	—	—	—	—

По формулам (2.5) находим параметры регрессии

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{13484 - 155,75 \cdot 85,58}{7492,25 - 85,58^2} = \frac{154,915}{168,31} = 0,92;$$

$$a = \bar{y} - b \cdot \bar{x} = 155,75 - 0,92 \cdot 85,58 = 77,02.$$

Получено уравнение регрессии:

$$y = 77,02 + 0,92 \cdot x.$$

Параметр регрессии позволяет сделать вывод, что с увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб. (или 92 коп.).

После нахождения уравнения регрессии заполняем столбцы 7–10 таблицы 2.3.

2. Тесноту линейной связи оценит коэффициент корреляции (2.6):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,92 \cdot \frac{12,97}{16,53} = 0,722;$$

Т.к. значение коэффициента корреляции больше 0,7, то это говорит о наличии весьма тесной линейной связи между признаками.

Коэффициент детерминации:

$$r_{xy}^2 = 0,521.$$

Это означает, что 52% вариации заработной платы (у) объясняется вариацией фактора x – среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации (2,7):

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{68,97}{12} = 5,75\%.$$

Качество построенной модели оценивается как хорошее, так как \bar{A} не превышает 10%.

3. Оценку статистической значимости уравнения регрессии в целом проведем с помощью F -критерия Фишера. Фактическое значение F -критерия по формуле (2.9) составит

$$F_{\text{факт}} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2) = \frac{0,521}{1 - 0,521} \cdot 10 = 10,88.$$

Табличное значение критерия при пятипроцентном уровне значимости и степенях свободы $k_1 = 1$ и $k_2 = 12 - 2 = 10$ составляет $F_{\text{табл}} = 4,96$. Так как $F_{\text{факт}} = 10,41 > F_{\text{табл}} = 4,96$, то уравнение регрессии признается статистически значимым.

Оценку статистической значимости параметров регрессии и корреляции проведем с помощью t -статистики Стьюдента и путем расчета доверительного интервала каждого из параметров.

Табличное значение t -критерия для числа степеней свободы $df = n - 2 = 12 - 2 = 10$ и уровня значимости $\alpha = 0,05$ составит $t_{\text{табл}} = 2,23$.

Определим стандартные ошибки m_a , m_b , $m_{r_{xy}}$ (остаточная дисперсия

на одну степень свободы $S_{\text{ост}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2} = \frac{1574,91}{10} = 157,49$):

$$m_a = \sqrt{S_{\text{ост}}^2 \frac{\sum x^2}{n^2 \sigma_x^2}} = \sqrt{157,49 \cdot \frac{89907}{12^2 \cdot 164,94}} = 24,42;$$

$$m_b = \sqrt{\frac{S_{\text{ост}}^2}{n \cdot \sigma_x^2}} = \sqrt{\frac{157,49}{12 \cdot 164,94}} = 0,282;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \sqrt{\frac{1 - 0,521}{12 - 2}} = 0,219.$$

Тогда

$$t_a = \frac{a}{m_a} = \frac{77,02}{24,42} = 3,15;$$

$$t_b = \frac{b}{m_b} = \frac{0,92}{0,282} = 3,26;$$

$$t_{r_{xy}} = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0,722}{0,219} = 3,30.$$

Фактические значения t -статистики превосходят табличное значение:

$$t_a = 3,26 > t_{\text{табл}} = 2,3; t_b = 3,16 > t_{\text{табл}} = 2,3; t_{r_{xy}} = 3,25 > t_{\text{табл}} = 2,3,$$

поэтому параметры a , b и r_{xy} не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительные интервалы для параметров регрессии a и b . Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a = t_{\text{табл}} \cdot m_a = 2,23 \cdot 24,42 = 54,46;$$

$$\Delta_b = t_{\text{табл}} \cdot m_b = 2,23 \cdot 0,282 = 0,63.$$

Доверительные интервалы

$$\gamma_a = a \pm \Delta_a = 77,02 \pm 54,46 \text{ и } 22,56 \leq a^* \leq 131,48;$$

$$\gamma_b = b \pm \Delta_b = 0,92 \pm 0,63 \text{ и } 0,29 \leq b^* \leq 1,55$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p = 1 - \alpha = 0,95$ параметры a и b , находясь в указанных границах, не принимают нулевых значений, т.е. являются статистически значимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит: $x_0 = \bar{x} \cdot 1,07 = 85,6 \cdot 1,07 = 91,6$ руб., тогда индивидуальное прогнозное значение заработной платы составит: $\hat{y}_0 = 77,02 + 0,92 \cdot 91,6 = 161,29$ руб.

5. Ошибка прогноза составит:

$$m_{\hat{y}_0} = \sqrt{S_{\text{ост}}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot \sigma_x^2} \right)} = \sqrt{157,49 \cdot \left(1 + \frac{1}{12} + \frac{(91,6 - 85,6)^2}{12 \cdot 164,94} \right)} = 13,17.$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$$\Delta_{\hat{y}_0} = t_{\text{табл}} \cdot m_{\hat{y}_0} = 2,23 \cdot 13,17 = 29,37.$$

Доверительный интервал прогноза:

$$\gamma_{\hat{y}_0} = \hat{y}_0 \pm \Delta_{\hat{y}_0} = 161,29 \pm 29,37 \text{ и } 131,92 \leq y_0^* \leq 190,66.$$

Выполненный прогноз среднемесячной заработной платы является надежным ($p = 1 - \alpha = 1 - 0,05 = 0,95$) и находится в пределах от 131,92 руб. до 190,66 руб.

6. В заключение решения задачи построим на одном графике исходные данные и теоретическую прямую (рис. 2.1):

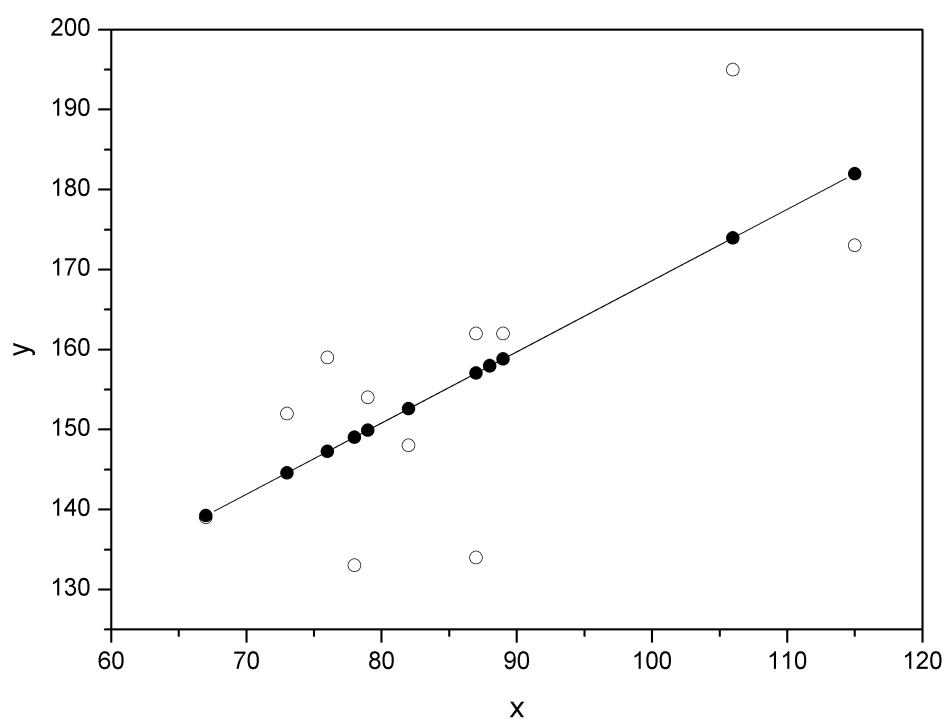


Рис. 2.1.

2.3. Решение типовой задачи в MS Excel

С помощью инструмента анализа данных **Регрессия** можно получить результаты регрессионной статистики, дисперсионного анализа, доверительных интервалов, остатки и графики подбора линии регрессии.

Если в меню сервис еще нет команды **Анализ данных**, то необходимо сделать следующее. В главном меню последовательно выбираем **Сервис**→**Надстройки** и устанавливаем «флажок» в строке **Пакет анализа** (рис. 2.2):

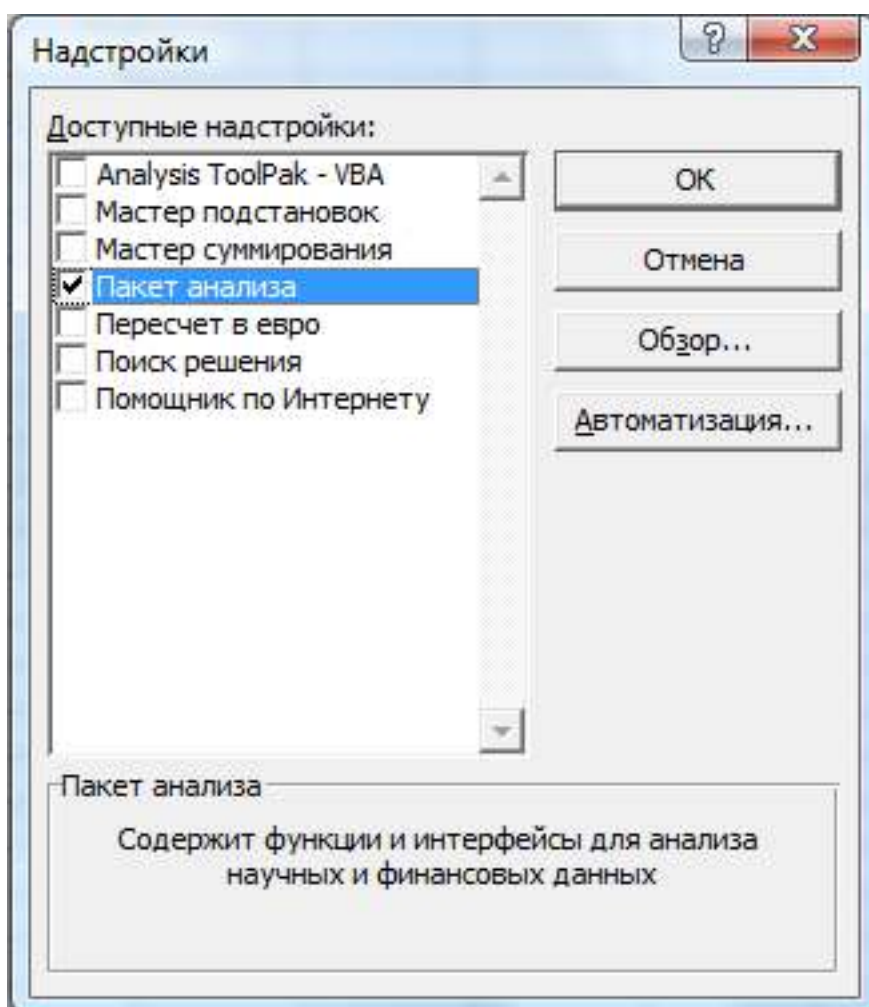


Рис. 2.2

Далее следуем по следующему плану.

1. Если исходные данные уже внесены, то выбираем **Сервис**→**Анализ данных**→**Регрессия**.

2. Заполняем диалоговое окно ввода данных и параметров вывода (рис. 2.3):

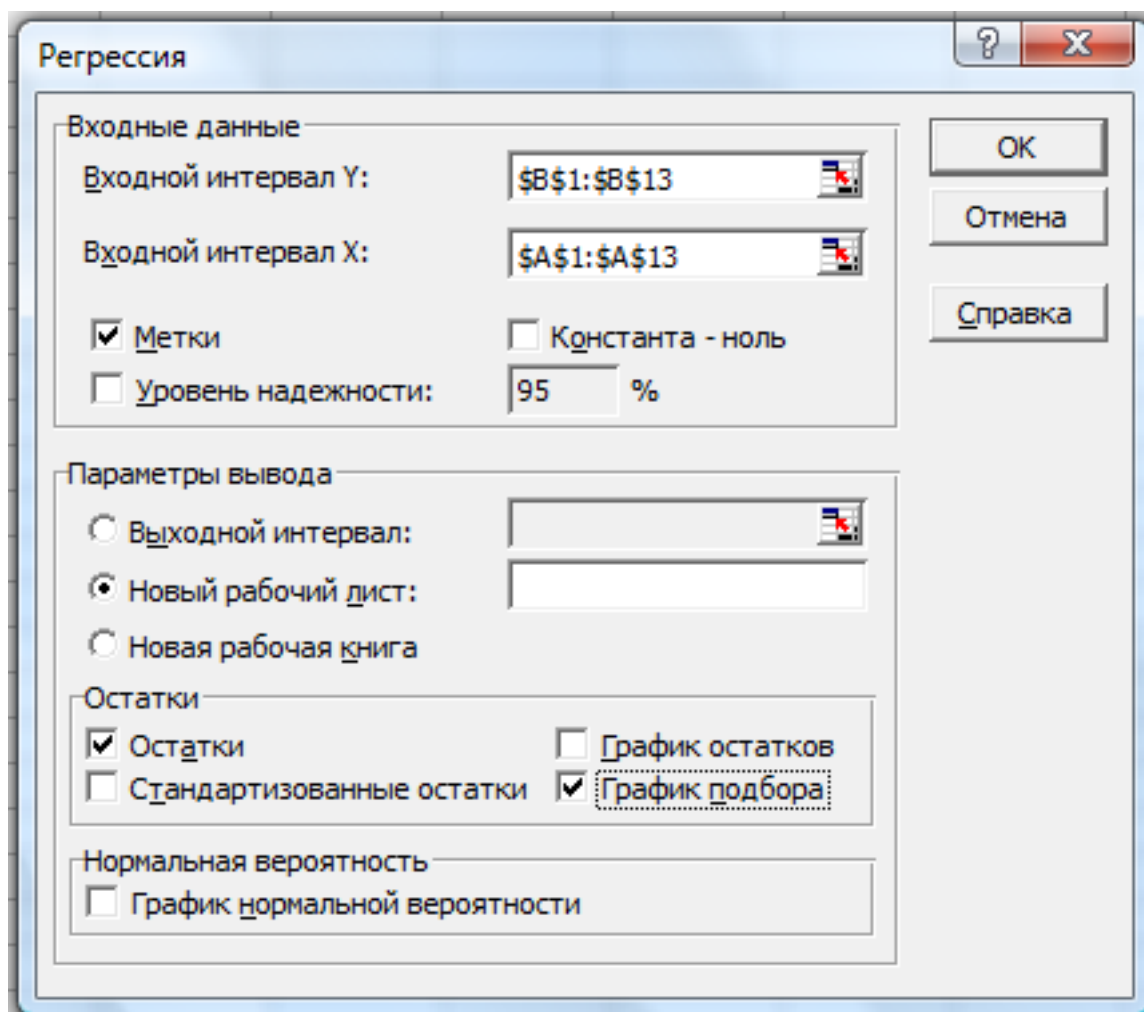


Рис. 2.3

Здесь:

Входной интервал Y – диапазон, содержащий данные результирующего признака;

Входной интервал X – диапазон, содержащий данные признака-фактора;

Метки – «флажок», который указывает, содержит ли первая строка названия столбцов;

Константа – ноль – «флажок», указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно указать произвольное имя нового листа (или не указывать, тогда результаты выводятся на вновь созданный лист).

Получаем следующие результаты для рассмотренного выше примера:

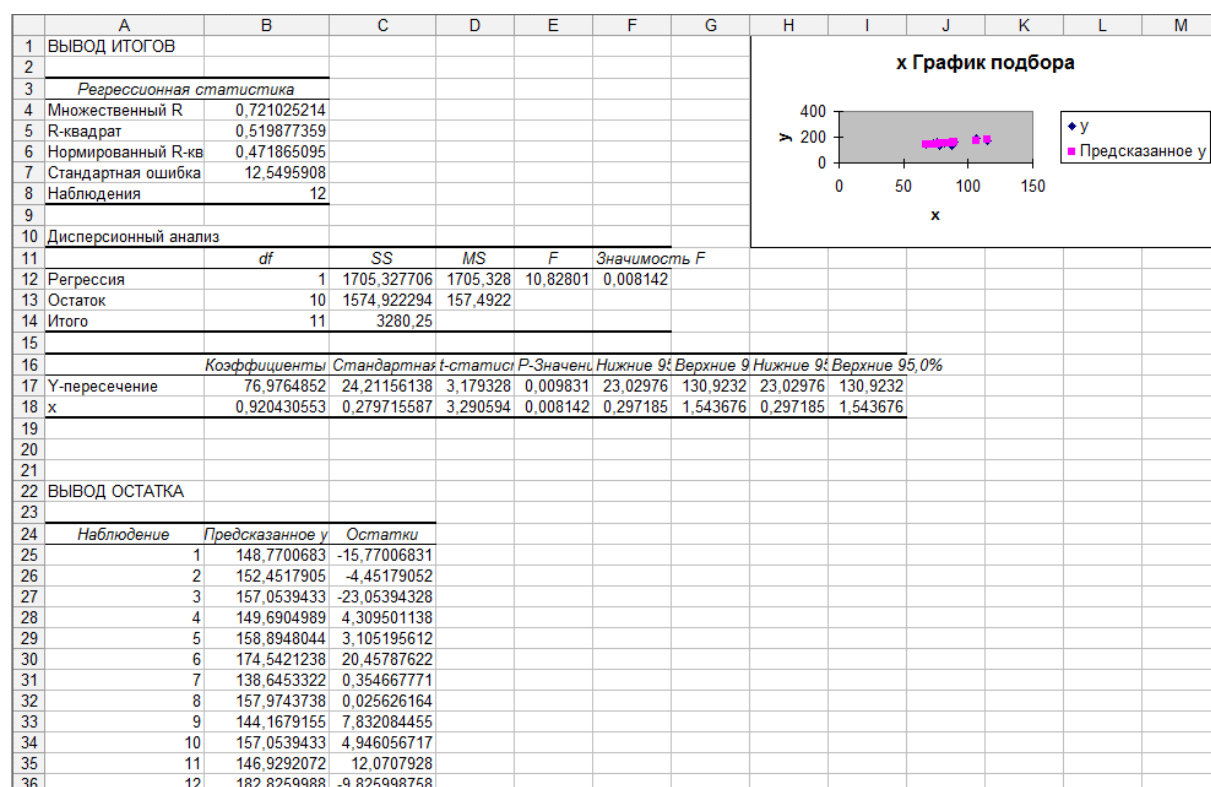


Рис. 2.4

Откуда выписываем, округляя до 4 знаков после запятой и переходя к нашим обозначениям:

Уравнение регрессии:

$$\hat{y}_x = 76,9765 + 0,9204x.$$

Коэффициент корреляции:

$$r_{xy} = 0,7210.$$

Коэффициент детерминации:

$$r_{xy}^2 = 0,5199.$$

Фактическое значение F -критерия Фишера:

$$F = 10,8280$$

Остаточная дисперсия на одну степень свободы:

$$S_{\text{ост}}^2 = 157,4922.$$

Корень квадратный из остаточной дисперсии (стандартная ошибка):

$$S_{\text{ост}} = 12,5496.$$

Стандартные ошибки для параметров регрессии:

$$m_a = 24,2116, \quad m_b = 0,2797.$$

Фактические значения t -критерия Стьюдента:

$$t_a = 3,1793, \quad t_b = 3,2906.$$

Доверительные интервалы:

$$23,0298 \leq a^* \leq 130,9232,$$

$$0,2972 \leq b^* \leq 1,5437.$$

Как видим, найдены все рассмотренные выше параметры и характеристики уравнения регрессии, за исключением средней ошибки аппроксимации (значение t -критерия Стьюдента для коэффициента корреляции совпадает с t_b). Результаты «ручного счета» от машинного отличаются незначительно (отличия связаны с ошибками округления).

Задания для контрольной работы

Задача 1. По территориям региона приводятся данные за 199X г. (p_1 – число букв в полном имени, p_2 – число букв в фамилии):

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	$78 + p_1$	$133 + p_2$
2	$80 + p_2$	148
3	87	$135 + p_1$
4	79	154
5	106	$157 + p_1$
6	$106 + p_1$	195
7	67	139
8	98	$158 + p_2$
9	$73 + p_2$	152
10	87	162
11	86	$146 + p_2$
12	$110 + p_1$	173

Требуется:

1. Построить линейное уравнение парной регрессии y по x .
2. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F -критерия Фишера и t -критерия Стьюдента.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x , составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

6. На одном графике отложить исходные данные и теоретическую прямую.

7. Проверить вычисления в **MS Excel**.

Рекомендации к выполнению контрольной работы

Практические задания по курсу «Эконометрика» следует выполнять в тетради или на листах бумаги формата А4 (листы скрепляются и заполняются с одной стороны). Работа обязательно должна содержать титульный лист с указанными на нем фамилии, полного имени и номера группы студента. Данные каждого варианта определяется параметрами p_1 , p_2 . При выполнении контрольных заданий студент должен подставить там, где это необходимо, вместо буквенных параметров индивидуальные анкетные характеристики: p_1 – число букв в полном имени студента; p_2 – число букв в фамилии студента.

Приложение 1.

Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

Приложение 2.

Критические значения t -критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний).

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	00,10	0,05	0,01		00,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Список литературы

Основная:

1. Эконометрика: Учебник / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2006. – 576 с.
2. Практикум по эконометрике: Учебн. пособие / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2006. – 344 с.
3. Эконометрика: Учебно-методическое пособие / Шалабанов А.К., Роганов Д.А. – Казань: ТИСБИ, 2004. – 198 с.
4. Доугерти К. Введение в эконометрику: Пер. с англ. – М.: ИНФРА-М, 1999. – 402 с.

Дополнительная:

5. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2002. – 311 с.
6. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. – М.: Дело, 2001. – 400 с.
7. Катышев П.К., Магнус Я.Р., Пересецкий А.А. Сборник задач к начальному курсу эконометрики. – М.: Дело, 2002. – 208 с.
8. Эконометрика: Учебник / Тихомиров Н.П., Дорохина Е.Ю. – М.: Издательство «Экзамен», 2003. – 512 с.
9. Сборник задач по эконометрике: Учебное пособие для студентов экономических вузов / Сост. Е.Ю. Дорохина, Л.Ф. Преснякова, Н.П. Тихомиров. – М.: Издательство «Экзамен», 2003. – 224 с.
10. Гмурман В.Е. Теория вероятностей и математическая статистика: Учебн. пособие для вузов. – М.: Высш. шк., 2002. – 479 с.